

---

# Le test de concordance comme outil d'évaluation en ligne du raisonnement des professionnels en situation d'incertitude

---

**Bernard Charlin**

Université de Montréal, CANADA  
[bernard.charlin@umontreal.ca](mailto:bernard.charlin@umontreal.ca)

**Driss Kazi-Tani**

Université de Montréal, CANADA  
[driss.kazi.tani@umontreal.ca](mailto:driss.kazi.tani@umontreal.ca)

**Robert Gagnon**

Université de Montréal, CANADA  
[robert.gagnon@umontreal.ca](mailto:robert.gagnon@umontreal.ca)

**Robert Thivierge**

Université de Montréal, CANADA  
[robert.thivierge@umontreal.ca](mailto:robert.thivierge@umontreal.ca)

## Compte rendu d'expérience

### Résumé

Une part significative de la compétence dans les professions relève de la capacité à composer avec l'incertitude. Le test de concordance est un outil d'évaluation standardisé de raisonnement en contexte d'incertitude qui peut être utilisé dans une variété de contextes d'utilisation. Cet article vise à décrire le format de l'outil, la méthode de correction et les moyens qui ont été mis en place pour administrer le test en ligne. Selon les résultats d'une série de travaux publiés à ce sujet, l'expérience acquise pousse à recommander l'utilisation du test de concordance dans toute profession qui implique de raisonner en contexte d'incertitude.

### Abstract

A significant part of professional competence stems from the ability to deal with uncertainty. The correlation test is a standardized assessment tool for reasoning in a context of uncertainty that may be put to a variety of uses. This article describes the test format, scoring method and means that have been developed to administer the test online. Based on a series of published studies, acquired experience suggests the usefulness of the correlation test in any profession involving reasoning in context of uncertainty.

### Introduction

Une part significative de la compétence dans les professions relève de la capacité à composer avec l'incertitude (Fox, 2000; Johnson, 1988). Les problèmes posés par certaines situations professionnelles peuvent être résolus par l'application simple de règles et de connaissances, mais les professionnels sont souvent confrontés à des situations complexes ou ambiguës qui demandent une réflexion pour interpréter les données et pour progresser vers la solution. La capacité à raisonner en contexte d'incertitude et à résoudre des problèmes mal définis caractérise l'exercice professionnel. Si les institutions d'enseignement forment souvent bien à l'utilisation des savoirs techniques qui permettent la résolution des problèmes bien définis, elles forment et évaluent difficilement le raisonnement en contexte d'incertitude (Schön, 1983).

Les connaissances nécessaires à la résolution des problèmes simples liés à des situations pour lesquelles il y a une seule bonne réponse peuvent être adéquatement et fidèlement évaluées

avec des outils standardisés tels que les questions à choix multiples à contexte riche (Case et Swanson, 2001). Il existe cependant un besoin de méthodes innovatrices permettant d'évaluer en contexte d'incertitude. Le test de concordance est basé sur une théorie cognitive, la théorie des scripts, qui met l'accent sur les réseaux de connaissances qu'utilisent les individus en situation de résolution de problème (Charlin, Tardif et Boshuizen, 2000). Elle assume que les individus traitent les données liées à la situation en activant des réseaux de connaissances préalables, organisés pour la résolution de tâches, les « scripts ». Le format du test (Charlin, Roy, Brailovsky et van der Vleuten, 2000) est fait de vignettes décrivant une situation problématique suivie par des propositions d'options possibles. Une nouvelle information est alors proposée et les candidats doivent décider si la nouvelle information augmente, diminue ou est sans influence sur la qualité de cette option. Les réponses sont captées sur une échelle de Likert, reflétant la façon dont l'information est traitée en situation de résolution de problème, selon la théorie des scripts (Charlin, Tardif *et al.*, 2000).

Les tests de concordance ont été jusqu'à maintenant utilisés en sciences de la santé (médecine, chirurgie, radiologie, sages-femmes, pharmacie, ergothérapie). Ils peuvent toutefois être utilisés dans toutes les professions qui gèrent de l'incertitude. L'idée essentielle est de placer le candidat devant une situation complexe requérant une tâche cognitive précise. Dans de telles situations, il existe une variabilité dans les réponses d'experts. Pour tenir compte de cette variabilité, le processus de score consiste à faire passer le test à un panel de référence et à donner au candidat un score qui reflète le nombre de membres du panel qui ont donné la même réponse. Ces principes permettent d'introduire des situations de vie réelle, avec toute leur complexité, dans les examens.

Le but de cet article est de présenter aux responsables de la formation dans les autres professions le format de l'outil, la méthode de correction, ainsi que les qualités psychométriques qui ont été établies à travers une série de travaux de recherche, tout en soulignant l'intérêt d'administrer et de corriger le test en ligne. Les publications concernant les études réalisées peuvent être téléchargées depuis le site <http://www.cme.umontreal.ca/tcs/>.

## Format du test

Pour chaque item, un cas est présenté, contenant soit insuffisamment de données pour résoudre le problème, soit une ambiguïté ou des conflits de valeurs (évaluation du raisonnement éthique par exemple), de sorte qu'un expert ne peut donner la solution au problème au vu du contenu de cette seule vignette et que diverses options sont pertinentes. Plusieurs questions sont alors posées. Chacune contient une option de résolution du cas, suivie par la présentation d'une nouvelle donnée. La tâche du candidat est de porter un jugement sur l'effet qu'a cette nouvelle donnée sur le statut de l'option proposée (Charlin, Roy *et al.*, 2000). Les questions suivantes, liées au même cas, explorent l'effet d'autres données sur d'autres options. La Figure 1, tirée des travaux de J.-P. Fournier (Fournier *et al.*, 2006), illustre le format du test.

## Méthode d'établissement des scores

On demande à des personnes expérimentées dans le domaine (enseignants ou praticiens) de remplir le test individuellement. Ces personnes constituent le panel de référence.

Vous êtes à l'urgence. Une patiente de 65 ans est référée par une clinique pour suspicion d'embolie pulmonaire. Elle a des antécédents de diabète, d'insuffisance cardiaque et de bronchopneumopathie chronique obstructive.

Si vous pensiez faire (option d'examen complémentaire)	Et qu'alors vous trouvez (nouvelle information obtenue par examen clinique ou examen complémentaire)	L'effet sur la nécessité de demander ce test est le suivant
Q1- Un angioscanner thoracique	Un traitement par metformine (Glucophage retard®)	-2 -1 0 +1 +2
Q2- Un dosage des D dimères	Un cancer de l'ovaire en cours de traitement	-2 -1 0 +1 +2
Q3- Un écho-Doppler veineux	Un signe de Homans	-2 -1 0 +1 +2

-2 : absolument contre-indiqué  
 -1 : peu utile ou plutôt néfaste  
 0 : non pertinent dans cette situation  
 +1 : utile et souhaitable  
 +2 : indispensable

Figure 1. Exemple d'item (un item étant un cas clinique avec les questions qui lui sont liées)

La distribution des réponses des membres du panel pour les 3 questions du cas présenté à la Figure 1 est montrée dans les cellules blanches. Les points reçus pour ces réponses par les candidats sont montrés dans les cellules grises.

Si vous pensiez faire	Et qu'alors vous trouvez	L'effet sur la nécessité de demander ce test est le suivant				
		-2	-1	0	+1	+2
Un angioscanner thoracique (Q1)	Un traitement par metformine (Glucophage retard®)	4	3	2	0	0
<b>Points attribués</b>		<b>1</b>	<b>0,75</b>	<b>0,5</b>	<b>0</b>	<b>0</b>
Un dosage des D dimères (Q2)	Un cancer de l'ovaire en cours de traitement	0	0	4	4	1
<b>Points attribués</b>		<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0,25</b>
Un écho-Doppler veineux (Q3)	Un signe de Homans	0	1	5	2	1
<b>Points attribués</b>		<b>0</b>	<b>0,2</b>	<b>1</b>	<b>0,4</b>	<b>0,2</b>

Figure 2. Création des clés de correction pour chaque question (panel de 9 personnes)

Chacune de leurs réponses contribue à la création de la grille de correction, en suivant la méthodologie décrite dans la littérature (Gagnon, Charlin, Coletti, Sauvé et van der Vleuten, 2005) : le nombre de réponses pour chaque question est enregistré. Les candidats reçoivent un crédit de points qui reflète le nombre de membres du panel qui ont répondu comme eux. Un point est attribué à la réponse modale à la question; les autres réponses donnent lieu à un crédit de points proportionnel à leur nombre. Les réponses non choisies reçoivent 0. Le processus consiste, pour chaque question, à diviser le nombre de réponses à chaque ancrage de l'échelle de Likert par la valeur modale pour cette question. Dans l'exemple des figures 1 et 2, si pour une question 4 membres du panel (parmi 9) ont choisi l'ancrage « -2 », cette réponse reçoit 1 point (4/4). Si 3 ont choisi « -1 », cette réponse reçoit 0,75 point (3/4) et si 2 membres ont choisi l'ancrage « 0 », cette réponse reçoit 0,5 (2/4). Le score total pour le test est la somme de tous les crédits obtenus pour chaque question. Les scores totaux sont ensuite transformés pour obtenir une note totale maximum de 100.

### Qualités psychométriques du test

Les études publiées (Charlin, Desaulniers, Gagnon, Blouin et van der Vleuten, 2002;

Charlin, Gagnon, Sauvé et Coletti, soumis; Charlin *et al.*, 2000; Charlin et van der Vleuten, 2004; Fournier *et al.*, 2006; Gagnon *et al.*, 2005) montrent qu'il faut une soixantaine de questions pour atteindre une bonne fidélité de mesure, avec une valeur du coefficient de cohérence interne de Cronbach atteignant ou dépassant 0,75. Le temps de passage pour un test de 60 questions est d'environ une heure. Selon les étudiants, les résidents et les professionnels, le test est intéressant à passer, car les tâches ressemblent à celles de la pratique réelle. Le test, qu'il soit présenté sur papier ou sur écran d'ordinateur, est relativement facile à administrer, requérant peu de ressources matérielles ou humaines. Le processus de correction est cependant complexe et requiert l'utilisation de programmes informatisés. De tels programmes sont disponibles sur le site <http://www.cme.umontreal.ca/tcs/html/correction.html>.

Un test qui mesure réellement les capacités de raisonnement professionnel devrait donner les scores les plus élevés aux personnes qui ont le plus d'expérience professionnelle (validité de construit). Or les tests usuels donnent des résultats paradoxaux : ce ne sont pas les professionnels expérimentés qui obtiennent les meilleurs scores, mais les personnes de niveau intermédiaire, les rési-

dents en fin de formation, proches de leurs examens de certification. Cet effet est nommé « effet intermédiaire » dans la littérature en éducation médicale (van der Vleuten et Newble, 1995). Pour le test de concordance, toutes les études (Charlin, Desaulniers *et al.*, 2002; Charlin *et al.*, soumis; Charlin *et al.*, 2000; Charlin et van der Vleuten, 2004; Fournier *et al.*, 2006; Gagnon *et al.*, 2005) ont montré une progression des scores depuis les novices (les étudiants) vers les médecins expérimentés, en passant par les différents niveaux de résidence. Cette disparition de l'effet intermédiaire a par exemple été démontrée à propos du raisonnement des radio-oncologues (Lambert, 2005) ou des médecins d'urgence (Carrière, 2005).

La validité prédictive du test a été montrée par une étude (Brailovsky, Charlin, Beausoleil, Coté et van der Vleuten, 2001) dans laquelle un test de concordance a été administré à tous les étudiants d'une faculté immédiatement avant leurs examens de fin de formation. Le test portait sur les connaissances cliniques en chirurgie. La cohorte d'étudiants qui s'est ensuite engagée en formation de médecine familiale a été suivie jusqu'à l'examen d'obtention de la spécialité, deux ans plus tard. Compte tenu de l'objet de mesure du test, l'organisation des connaissances adaptées au

raisonnement clinique, l'hypothèse était que le test prédirait bien les résultats des tests de raisonnement clinique et moins bien ceux concernant les habiletés cliniques, mesurées par un examen clinique objectif et structuré (ECOS). Les résultats ont confirmé cette hypothèse, suggérant que les étudiants qui organisent bien leurs connaissances à un moment de leur formation continuent de le faire aux stades ultérieurs de formation.

Le processus de score, composante clé des tests de concordance, incorpore la variabilité que les experts démontrent lorsqu'ils sont confrontés aux problèmes complexes et mal définis qui caractérisent la vie professionnelle. Cette variabilité soulève des questions à propos du nombre de membres requis sur le panel pour obtenir des scores stables, de l'influence de la composition du panel sur les scores ou du degré de variabilité des réponses parmi les experts.

Il a été démontré (Gagnon *et al.*, 2005) que pour obtenir la stabilité de scores indispensables pour des examens à haut enjeu (examens de certification), le panel doit comporter au moins une quinzaine de membres. Pour des examens à enjeu moindre, 8 à 10 membres suffisent. L'expérience acquise dans d'autres types d'examen peut faire paraître difficile l'atteinte de tels chiffres. En matière de test de concordance, il s'agit simplement de demander aux membres pressentis de compléter individuellement le test exactement comme des étudiants le feraient. Leurs réponses sont ensuite compilées pour bâtir la grille de correction. Ce processus peu exigeant en temps est très bien accepté.

Une étude menée en médecine générale (Charlin *et al.*, soumis) a comparé les scores de résidents obtenus à partir de deux panels, l'un composé d'enseignants, l'autre de médecins de pratique privée sans fonc-

tion d'enseignement, mais suivant régulièrement des sessions de formation continue. Le rang des résidents était le même (corrélation à 0,98). Par contre, les scores étaient un peu plus élevés lorsque la correction utilisait le panel des enseignants. Il a par ailleurs été démontré que les questions qui discriminent bien en fonction du degré d'expérience sont celles pour lesquelles il existe une variabilité entre les réponses des membres du panel. En effet, si tous les membres du panel donnent la même réponse, la question devient similaire à une question à choix multiple. Par contre, si les réponses divergent totalement, il s'agit vraisemblablement d'une question mal construite.

### **Développements informatiques réalisés pour permettre la conception et l'administration en ligne des tests de concordance**

**Besoins :** L'application a été conçue pour répondre à trois besoins :

1. Faciliter toutes les étapes de réalisation d'un test de concordance, soit la construction, la collecte des réponses des participants et la compilation des résultats.
2. Économiser du temps, tant aux constructeurs du test qu'aux personnes évaluées.
3. Enrichir l'expérience collective concernant la mise en pratique du test de concordance à travers la rétroaction des utilisateurs.

**Conception :** Dans un but de souplesse d'utilisation, la solution apportée devait également se garder de contraindre les constructeurs de tests à reprendre le système de bout en bout lorsqu'ils veulent travailler sur le test. Cela a amené à structurer chaque étape dans un module indépendant; le constructeur peut ainsi utiliser les services d'un module puis télécharger son travail sur son ordinateur pour continuer selon la méthode « papier-crayon »

ou poursuivre le processus avec le module suivant (voir la Figure 3).

**Réalisation :** À ce jour, deux systèmes ont été construits et sont disponibles en ligne (voir la Figure 3). Ils sont compatibles avec les navigateurs gratuits Explorer 5.5 et Mozilla Firefox.

1. Le test de concordance dans sa version texte originale comprend trois modules utilisables isolément ou d'une manière intégrée :
  - a. Le « module du constructeur ». Il permet aux experts en contenu de bâtir un cas de test de concordance de script (TCS) en ligne et de le sauvegarder structuré dans son ordinateur (fichier ZIP contenant des dossiers et les fichiers associés).
  - b. Le « module du gestionnaire du test ». Il permet de faire passer un test construit avec le « module du constructeur » à des participants (membres du panel de référence et candidats à évaluer). Les réponses enregistrées peuvent être téléchargées pour un traitement personnalisé (format compatible MS-Excel ou SPSS) ou pour être confiées au « module du correcteur », qui compilera les résultats.
  - c. Le « module du correcteur ». Il peut accepter trois types d'entrées de données : un fichier de données enregistré avec le « module du gestionnaire du test », un fichier Excel respectant un certain format ou des données entrées manuellement. La compilation et les données initiales peuvent être téléchargées pour une sauvegarde ou d'autres traitements à l'aide de MS-Excel ou SPSS.
2. Le test de concordance appliqué aux domaines de diagnostic visuel (radiologie, dermatologie, etc.) a la même architecture de base que la version texte présentée précédemment, mais avec la particularité de présenter des images réactives aux clics de souris des usa-

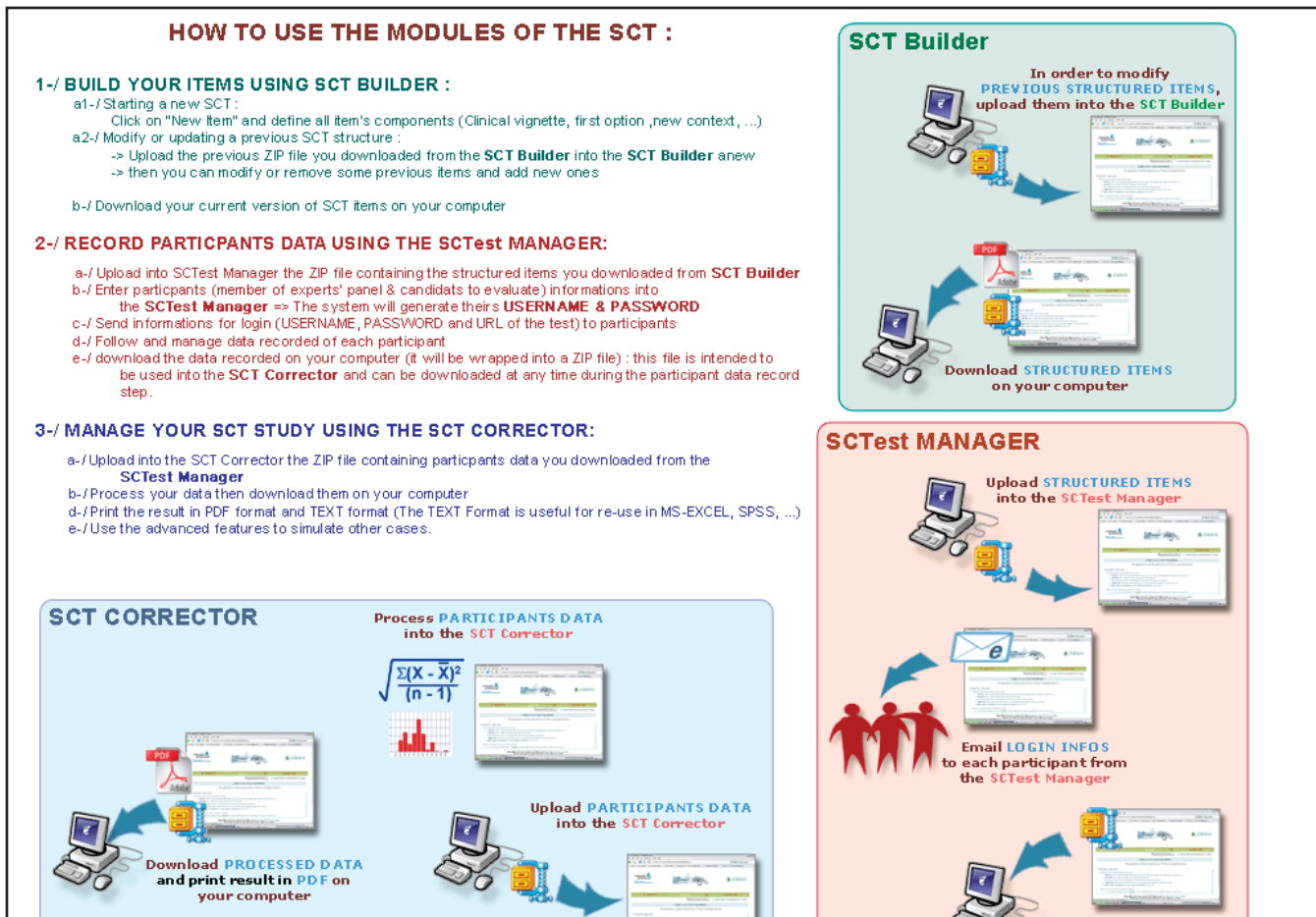


Figure 3. Modules du test de concordance

gers; les coordonnées des clics enregistrées sont comparées à celles du panel d'experts pour déterminer si l'anomalie a bien été détectée. Le constructeur du test peut au besoin adapter le comportement du « module du correcteur » aux contenus pédagogiques et aux candidats à évaluer en paramétrant manuellement certaines variables du système (précision des clics de souris, délais de réponse, etc.). Une version de démonstration de cette technologie est accessible sur le site <http://www.cme.umontreal.ca/sctimages/>.

### Administration du test en ligne et application à de nouveaux champs d'évaluation

Sibert, Darmoni, Dahamna, Weber et Charlin (2005) ont récemment pré-

senté une expérience d'administration d'un test de concordance en ligne pour tous les urologues français. Leur panel comportait 26 membres et la population testée était composée de 100 urologues, 100 résidents et 50 étudiants. Tous les participants remplissaient le test à partir d'un site Web. L'intention des auteurs est d'aider les urologues à maintenir leur qualité de raisonnement clinique en se testant à partir de leur lieu de pratique ou de leur domicile. Une expérience similaire est menée auprès des neurochirurgiens français (Caire, Sol, Moreau, Isidori et Charlin, 2004). Le test est également utilisé en ligne par l'Université de Montréal pour aider les médecins en exercice à déceler leur besoin de formation à partir de leurs résultats au test. La non-concordance avec des questions permet de déceler les zones sur lesquelles

doit porter la formation. Des exemples sont fournis sur le site <http://www.cme.umontreal.ca/html/TCS.html>.

Le test de concordance, conçu pour exposer les participants à des tâches de la vie professionnelle réelle dans toute leur complexité et vérifier dans quelle mesure leurs décisions sont conformes à celles prises par les membres d'un panel de référence, permet d'évaluer des domaines qui étaient jusqu'à maintenant difficilement évaluables de façon standardisée et objective. Citons par exemple le raisonnement préopératoire des chirurgiens (Meterissian, Zaboltyn, Gagnon et Charlin, soumis), la qualité du jugement éthique (Llorca, 2003) ou les capacités de perception et d'interprétation en radiologie : <http://www.cme.umontreal.ca/sctimages/>.

## Conclusion

Le test de concordance permet de confronter les participants à des situations complexes reflétant la réalité de la pratique professionnelle. Son processus de score offre des résultats à la fois valides et fiables. Il s'administre en ligne, fonctionne aussi bien en formation initiale que continue et permet de mesurer des capacités qui jusqu'à maintenant n'étaient pas évaluables de façon standardisée avec les outils actuels. L'expérience acquise dans les sciences de la santé pousse à recommander son utilisation dans toute autre profession qui implique de raisonner en contexte d'incertitude.

## Références

- Brailovsky, C., Charlin, B., Beausoleil, S., Coté, S. et van der Vleuten, C. (2001). Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: An exploratory study on the script concordance test. *Medical Education*, 35, 430-436.
- Caire, F., Sol, J. C., Moreau, J. J., Isidori, P. et Charlin, B. (2004). Autoévaluation des internes en neurochirurgie par tests de concordance de script (TCS) : le processus d'élaboration des tests. *Neurochirurgie*, 50, 66-72.
- Carrière, B. (2005). *Development and initial validation of a script concordance test for residents in a pediatric emergency medicine rotation*. Mémoire de maîtrise non publié, University of Illinois at Chicago.
- Case, S. M. et Swanson, D. B. (2001). *Constructing written test questions for basic and clinical sciences* (3<sup>e</sup> éd.). Philadelphia: The National Board of Medical Examiners.
- Charlin, B., Desaulniers, M., Gagnon, R., Blouin, D. et van der Vleuten, C. (2002). Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. *Teach Learn Med*, 14(3), 150-156.
- Charlin, B., Gagnon, R., Sauvé, E. et Coletti, M. (soumis). Composition of the panel of reference for concordance tests: Do teaching functions have an impact on examinees' ranks and absolute scores? *Medical Teacher*.
- Charlin, B., Roy, L., Brailovsky, C. et van der Vleuten, C. (2000). The script concordance test, a tool to assess the reflective clinician. *Teaching and Learning in Medicine*, 12, 189-195.
- Charlin, B., Tardif, J. et Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine*, 75, 182-190.
- Charlin, B. et van der Vleuten, C. (2004). Standardized assessment of reasoning in contexts of uncertainty: The script concordance approach. *Eval Health Prof*, 27(3), 304-319.
- Fournier, J.-P., Thiercelin, D., Pulcini, C., Alunni-Perret, V., Gilbert, E., Minguet, J.-M. et Bertrand, F. (2006). Évaluation du raisonnement clinique en médecine d'urgence : les tests de concordance des scripts décèlent mieux l'expérience clinique que les questions à choix multiples à contexte riche. *Pédagogie médicale*, 7, 1, 20-30. Récupéré le 18 avril 2006 du site de la revue, [http://www.pedagogie-medicale.org/vol71\\_recherche.pdf](http://www.pedagogie-medicale.org/vol71_recherche.pdf)
- Fox, R. (2000). Medical uncertainty revisited. Dans G. Albrecht, R. Fitzpatrick et S. Scrimshaw (dir.), *Handbook of social studies in health and medicine* (p. 409-425). London: Sage Publications.
- Gagnon, R., Charlin, B., Coletti, M., Sauvé, E. et van der Vleuten, C. (2005). Assessment in the context of uncertainty: How many members are needed on the panel of reference of a script concordance test? *Medical Education*, 39, 284-291.
- Johnson, E. (1988). Expertise and decision under uncertainty: Performance and process. Dans M. Chi, R. Glaser et M. Farr (dir.), *The nature of expertise* (p. 209-228). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lambert, C. (2005). *Le test de concordance de script: étude de validation d'un nouvel outil d'évaluation du raisonnement clinique des résidents en radio-oncologie*. Mémoire de maîtrise non publié, Université de Montréal. Récupéré le 18 avril 2006 du site Formation professionnelle continue de la Faculté de médecine de l'Université de Montréal, [http://www.cme.umontreal.ca/tcs/shared/LambertC\\_Memoire\\_maitrise.pdf](http://www.cme.umontreal.ca/tcs/shared/LambertC_Memoire_maitrise.pdf)
- Llorca, G., Roy, P. et Riche, B. (2003). Évaluation de résolution de problèmes mal définis en éthique clinique : variation des scores selon les méthodes de correction et selon les caractéristiques des jurys. *Pédagogie médicale*, 4, 2, 80-88. Récupéré le 18 avril 2006 du site de la revue, <http://www.pedagogie-medicale.org/4-2-llorca.pdf>
- Meterissian, S., Zabolotny, B., Gagnon, R. et Charlin, B. (soumis). Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *Am J Surg*.
- Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Sibert, L., Darmoni, S. J., Dahamna, B., Weber, J. et Charlin, B. (2005, 20 juin). Online clinical reasoning assessment with the script concordance test: A feasibility study. *BioMed Central Medical Informatics and Decision Making*, 5, 18-28. Récupéré le 18 avril 2006 du site de la revue, <http://www.biomedcentral.com/1472-6947/5/18>
- Van der Vleuten, C. P. et Newble, D. I. (1995). How can we test clinical reasoning? *Lancet*, 345(8956), 1032-1034.